

Penerapan K-Means untuk *Clustering* data stunting pada Kecamatan Kota Waingapu

Application of K-Means for Clustering stunting data in Waingapu City District

Alfrian C Talakua¹, Erwianta Gustial Radjah², Novem B Uly³

Program Studi Sistem Informasi, Universitas Kristen Wira Wacana Sumba

E-mail: ¹alfriantalakua@unkriswina.ac.id, ²erwiantaradjah@unrkriswina.ac.id, ³novemuly@unrkriswina.ac.id

KEYWORDS:

Stunting, K-means, Data mining

ABSTRACT

Stunting in Waingapu is an important issue for health services, the data collection and data processing process is one of the main factors in making decisions for the government in carrying out work programs, the shortage of health workers is the reason for processing data quickly and being archived well. This research is Waingapu which is one of the reasons one sub-district that contributes the highest stunting rate and needs to carry out data processing to determine future plans, data is available from 2021-2023, K - means is used to process the data to help group data clusters accurately. The K-mean results from 171 data produced 3 clusters with a Davies Bouldin value = 0.968 with the combination of data on poor nutrition for 53 children, adequate nutrition for 48 children and good nutrition for 70 children.

KATA KUNCI:

Stunting, K-means, Data mining

ABSTRAK

Stunting di Waingapu menjadi isu penting bagi layanan Kesehatan, proses pendataan dan pengolahan data menjadi salah satu faktor utama dalam pengambilan keputusan bagi pemerintah dalam menjalankan program kerja, kekurangan tenaga Kesehatan menjadi alasan untuk memproses data dengan cepat dan terarsip dengan baik Penelitian ini waingapu yang menjadi salah satu kecamatan penyumbang angka stunting terbanyak dan perlu melakukan pengolahan data untuk menentukan rencana mendatang, data yang tersedia dari 2021- 2023, K – means digunakan untuk mengolah data tersebut untuk membantu mengelompokan data cluster secara akurat. Hasil K-mean dari 171 data menghasilkan 3 cluster dengan Nilai Davies Bouldin = 0.968 dengan perangkian data Gizi buruk 53 anak, Gizi Cukup 48 anak dan Gizi baik 70 anak.

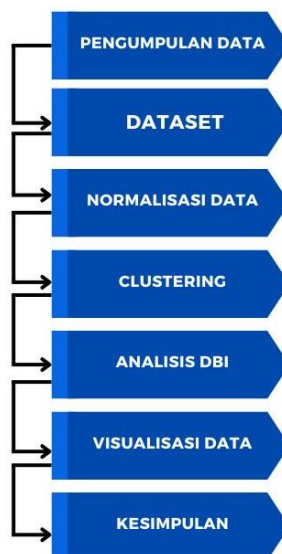
PENDAHULUAN

Stunting merupakan kondisi gagal pertumbuhan atau kondisi gizi buruk pada anak yang berpengaruh pada pertumbuhan tubuh dan otak yang diakibatkan oleh kekurangan gizi dalam waktu yang lama, dampaknya anak memiliki tinggi badan lebih pendek dari anak-anak seusianya juga memiliki berat badan yang tidak ideal. Kondisi ini mengakibatkan keterbatasan dalam berpikir, juga berpengaruh pada kemampuan kognitif dan fisik dalam tumbuh kembang yang dapat mengganggu kualitas hidup anak[1]. Berdasarkan data Kementerian Kesehatan presentase angka stunting di Indonesia pada tahun 2023 sebesar 21,5% hanya mengalami penurunan 1% dari tahun sebelumnya. Pemerintah menargetkan presentase data stunting di Indonesia harus mencapai 14 %. Segala upaya telah dilakukan untuk menekan angka stunting dengan berbagai program dari pemerintah, namun belum mendapatkan hasil yang maksimal. Terdapat 10 (sepuluh) provinsi penyumbang stunting tertinggi yaitu Sulawesi Barat, Nusa Tenggara Barat, Gorontalo, Aceh, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Barat,

Sulawesi Tenggara, Sulawesi Tengah, dan Nusa Tenggara Timur. Sumba timur merupakan salah satu daerah di provinsi nusa tenggara timur yang memiliki presentase angka stunting tinggi dengan prevalensi 11, 8 % pada tahun 2023. Laporan data stunting yang dikalkulasikan oleh petugas kesehatan menggunakan pengukuran tinggi dan berat badan untuk mendapatkan status gizi bayi sesuai dengan perhitungan dari WHO (*World Health Organization*)[2]. Kondisi geografis dan kekurangan tenaga kesehatan di sumba timur mengakibatkan proses pendataan dan pelaporan data cukup terhambat. Hasil dari pendataan yang dilakukan diklasifikasikan dalam gizi buruk dan gizi baik, namun beberapa variable perlu diambil dalam proses perhitungan untuk mendapatkan pemetaan data yang lebih baik. Jumlah data yang banyak dan proses pengolahan hingga visualisasi mempengaruhi hasil dari laporan presentase data stunting, disisi lain efektivitas dan ketepatan data laporan stunting menjadi hal yang sangat penting dalam pengambilan keputusan bagi pemerintah untuk penyusunan program kerja. Berdasarkan hal tersebut perlu menerapkan metode pengolahan data yang lain untuk menemukan pola atau hasil analisa data yang efektif yang disesuaikan dengan variable data stunting. Metode Clustering menjadi salah satu pilihan yang dapat digunakan untuk memetakan data yang disesuaikan dengan kemiripan nilai variable sehingga data dapat dibagi menjadi beberapa cluster untuk melihat hasil kelompok data yang bukan hanya pada gizi buruk dan gizi baik namun dapat membagi dalam beberapa cluster sehingga dapat diketahui variable yang paling berpengaruh dan tingkatan cluster hal ini akan membantu dalam memberikan rekomendasi dalam pengambilan keputusan perancangan program stunting. Penelitian ini akan dilakukan proses clustering data menggunakan metode K-Means, salah satu metode data mining yang digunakan untuk membagi data dalam beberapa cluster berdasarkan kemiripan data dengan nilai centroid yang ditetapkan sebagai data awal. Hasil Cluster akan dilihat data sebaiknya dibagi menjadi berapa kelompok terbaik. Berdasarkan kelompok tersebut dapat dilakukan analisis lanjut dari objek data yang telah dikelompokan[3].

METODE PENELITIAN

Penyusunan kerangka penelitian dibagi menjadi beberapa tahapan untuk mencapai hasil penelitian. Tahapan penelitian digambarkan pada gambar 1, tahapan penelitian menggunakan pendekatan metode K-means.



Gambar 1. Tahapan Penelitian

2.1 Pengumpulan Data Mahasiswa

Pada tahap ini dilakukan pengumpulan data balita pada puskesmas dan posyandu tahun 2021- 2023. Data dikumpulkan berdasarkan hasil pemeriksaan layanan. Berdasarkan data yang dikumpulkan pada tahun 2021 terdapat 51 balita, tahun 2022 terdapat 61 balita dan tahun 2023 terdapat 88 balita, terdapat penambahan balita disetiap tahunnya. Namun untuk melakukan clutstering tidak semua field data akan digunakan dalam proses analisis, sehingga data akan diseleksi untuk dijadikan dataset penelitian.

2.2 Dataset

Pada tahap ini proses mempersiapkan dataset yang diperlukan untuk melakukan clustering, proses ini dilakukan untuk seleksi data dengan field yang dibutuhkan pada, field yang dibutuhkan ialah Nama anak, Berat badan, tinggi badan dan lingkaran kepala yang dapat terlihat pada tabel 1. Total data yang digunakan pada penelitian ini sebanyak 171 data dari jumlah keseluruhan 200 data, 25 data dieliminasi karena tidak lengkap.

Tabel 1. Dataset Balita

No.	Nama Anak	Berat Badan	Tinggi badan	Lingkar Kepala
1.	Ag***** Fr***	11.2	96.5	15
2,	Agus**** A S***	12.5	93.8	14.5
3.	Anas*** K***	11	86	14
4.	Ant** N**** H***	10.6	83	0
5.	Ax** U*** M	9.5	78	15
6.	Ayunensi K. Kahi	12	94	15
7.	Desa**** M M***	8.8	78	14.5
8.	Dest***** J**	9.8	75.5	15.5
9.	Echa Rambu Pamar	13.4	90.5	15.5
10	FER****	12	90	16
11	F*** TIM****	8.9	75	14
12	FRAN**** M.A.D	11	89	14.5
13	GER**** U L	9	76	13
14	GI**** P.N	8.2	65.9	14
15	IN*** T*** A	11	85	14
16	I AR** ND***	6	64	13
17	LEONARD ADVENU	10	85	13
18	MARTINUS LILI	14	98	14
19	MICE*** A**** E	7	73.5	12.6
20	MIR****	13.5	94	15.5
...
171	YEH*****	8.1	85.5	11.5

2.3 Normalisasi data

Pada dataset yang telah dibuat tidak dapat langsung dianalisis karena perbedaan nilai yang cukup jauh antara besaran Tinggi badan dengan berat badan maupun lingkaran kepala, sehingga diperlukan tahapan normalisasi

data untuk sehingga mengurangi perbedaan besaran angka pada variable. Rumus Normalisasi diuraikan sebagai berikut;

$$\text{Nilai Normalisasi} = \frac{(\text{Nilai Awal} - \text{Nilai Minimum})}{(\text{Nilai Maksimal} - \text{Nilai Minimum})}$$

Gambar 2. Rumus Normaliasi[4]

Sehingga hasil normalisasi data dapat terlihat pada tabel berikut :

Tabel 2. Normalisasi Data

No.	Nama Anak	Berat Badan	Tinggi badan	Lingkar Kepala
1.	Ag***** Fr***	0,61	0,94	0,73
2.	Agus**** A S***	0,73	0,86	0,64
3.	Anas*** K***	0,59	0,65	0,55
4.	Ant** N**** H***	0,45	0,43	0,73
5.	Ax** U*** M	0,69	0,87	0,73
6.	Ayunensi K. Kahi	0,38	0,43	0,64
7.	Desa**** M M***	0,48	0,37	0,82
8.	Dest***** J**	0,82	0,77	0,82
9.	Echa Rambu Pamar	0,79	0,76	0,73
10.	FER****	0,69	0,76	0,91
11.	F*** TIM****	0,39	0,35	0,55
12.	FRAN**** M.A.D	0,59	0,73	0,64
13.	GER**** U L	0,40	0,38	0,36
14.	GI**** P.N	0,32	0,11	0,55
15.	IN*** T*** A	0,59	0,63	0,55
16.	I AR** ND***	0,11	0,05	0,36
17.	LEONARD ADVENU	0,50	0,63	0,36
18.	MARTINUS LILI	0,88	0,98	0,55
19.	MICE*** A**** E	0,21	0,31	0,29
20.	MIR****	0,83	0,87	0,82
...
171.	YEH*****	0.64	0,31	0,09

Hasil normalisasi data sudah menyeimbangkan besaran data setiap variabel sehingga proses clustering menggunakan K-Means dapat dilakukan secara optimal.

2.3 Clustering data

Pada tahap ini akan dilakukan clustering data menggunakan algoritma K-Means dengan beberapa Langkah sebagai berikut ;

Langkah-langkah melakukan clustering dengan metode K-Means adalah sebagai berikut:

- Pilih jumlah cluster k .
- Inisialisasi k pusat cluster ini bisa dilakukan dengan berbagai cara. Namun yang paling sering dilakukan adalah dengan cara random. Pusat-pusat cluster diberi nilai awal dengan angka-angka random.
- Alokasikan semua data/ objek ke cluster terdekat.

$$D(i, j) = \sqrt{(C_{1i} - C_{1j})^2 + (C_{2i} - C_{2j})^2 + \dots + (C_{ki} - C_{kj})^2}$$

Keterangan:

Dimana:

$D(i, j)$ = jarak data ke i ke pusat cluster j

C_{ki} = Data ke i pada atribut data ke k

C_{kj} = Data ke j pada atribut data ke k

- Hitung kembali pusat cluster dengan keanggotaan cluster yang sekarang. Pusat cluster adalah rata-rata dari semua data/ objek dalam cluster tertentu. Jika dikehendaki bisa juga menggunakan median dari cluster tersebut. Jadi rata-rata (mean) bukan satu-satunya ukuran yang bisa dipakai.
- Tugaskan lagi setiap objek memakai pusat cluster yang baru. Jika pusat cluster tidak berubah lagi maka proses clustering selesai. Atau, kembali ke langkah nomor 3 sampai pusat cluster tidak berubah lagi.

Berdasarkan tahapan diatas maka akan ditentukan nilai K yang digunakan untuk clustering ialah 3, 4 dan 5 sehingga proses pengujian akan membagi cluster dengan jumlah maksimal yaitu 5 cluster. Setelah itu akan dianalisis cluster yang baik berdasarkan nilai *Davis Boilden* dengan besaran angka yang mendekati angka 0. Proses Pengolahan data akan menggunakan software rapidminer sehingga dapat mempermudah dalam melakukan analisis dan visualisasi. [5]

2.4 Analisis Data

Pada tahap ini dilakukan proses analisis data untuk melihat jumlah data pada setiap cluster dan diurutkan berdasarkan rangking yang terbaik sehingga bisa terlihat urutan rangking dari setiap cluster yang dibuat. Hasil analisis akan membagi data stunting pada jumlah cluster terbaik.

2.5 Visualisasi Data

Pada tahap ini dilakukan visualisasi data untuk menampilkan data dalam bentuk tabel atau diagram sehingga dapat terlihat jumlah data yang tercluster dan jarak dari setiap data anak pada masing-masing cluster yang terbentuk.

HASIL DAN PEMBAHASAN

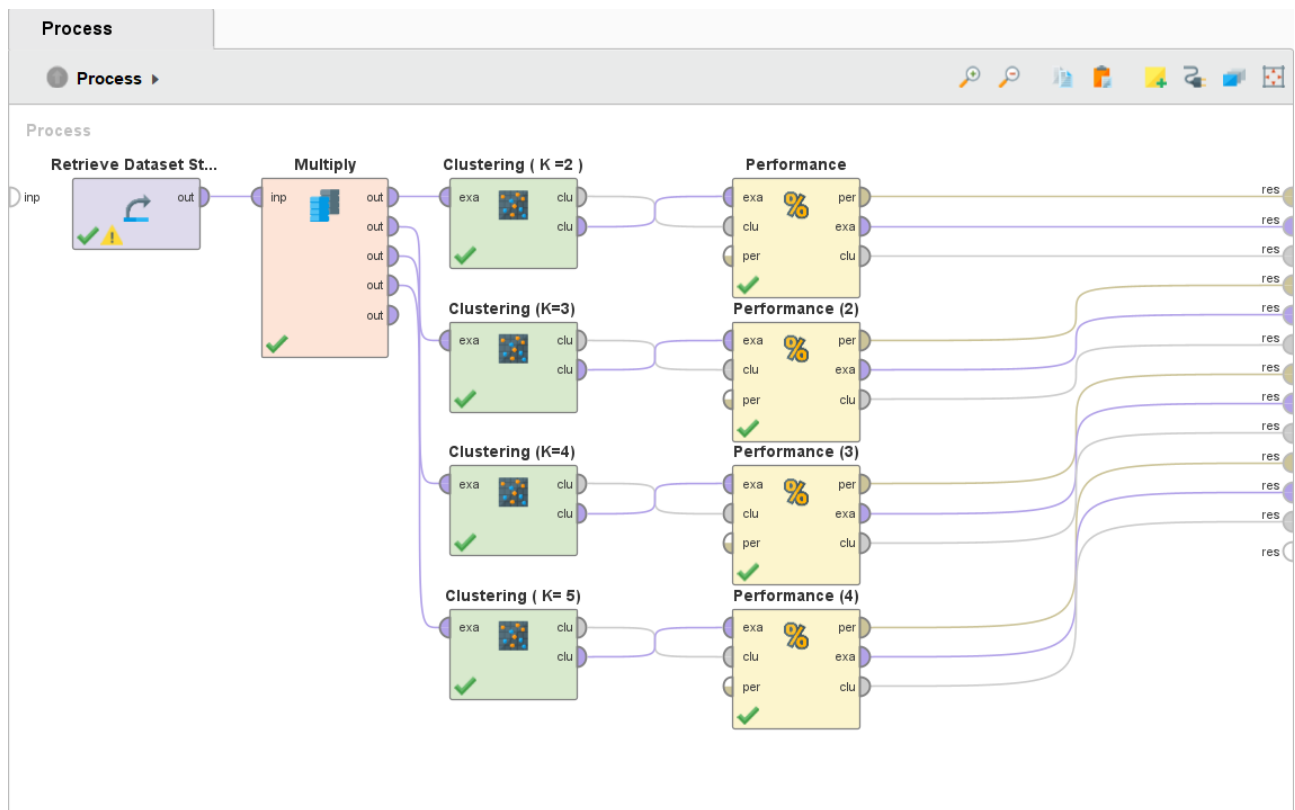
Proses pengolahan data dimulai dengan menginputkan dataset pada software Rapidminer, dataset diatur dengan mencantumkan nama anak sebagai Label, Berat badan, Tinggi badan dan Lingkar kepala yang memiliki tipe data interger.

Row No.	Na...	Berat Badan	Tinggi Badan	Lingkar Kep...
1	AGN...	0.610	0.938	0.727
2	agust...	0.733	0.864	0.636
3	anast...	0.590	0.652	0.545
4	AXEL...	0.448	0.435	0.727
5	ayun...	0.686	0.870	0.727
6	DES...	0.381	0.435	0.636
7	DES...	0.476	0.367	0.818
8	ECH...	0.819	0.774	0.818
9	faren...	0.790	0.761	0.727
10	FER...	0.686	0.761	0.909

Gambar 3. Input Dataset anak

Setelah proses input dataset proses selanjutnya akan memilih modul K-Mean dilanjutkan menentukan Nilai K sebagai nilai centroid untuk membagi cluster ialah 2 dengan measure types menggunakan Bregmen divergences dikarenakan data yang diolah semuanya merupakan data numeric, kemudian menggunakan maximum perpindahan titik centroid maximum 100 kali, sebelum dilakukan cluster harus menambahkan pengukuran performance untuk melihat nilai pembagian cluster.

Proses clustering akan dibagi dalam beberapa Nilai K = 2,3,4,5 dengan menggunakan modul multiply data, untuk membagi data sesuai dengan nilai K modul berikutnya menggunakan *Cluster Distance Performance* untuk mengukur performa jarak/distance Nilai K dalam bentuk nilai Davies Bouldin.



Gambar 4. Clusterisasi K-Means

Hasil dari Clusterisasi yang dilakukan menggunakan *Rapid Miner* data dibagi dalam cluster sesuai dengan nilai K. Data dengan **Nilai K = 2**; terbagi Cluster 0 = 83 items Cluster 1 = 88 items, **Nilai K = 3**; terbagi Cluster 0 = 53 items, Cluster 1 = 48 items, Cluster 2 = 70 items **Nilai K = 4** terbagi Cluster 0 = 56 items Cluster 1 = 43 items Cluster 2 = 48 Cluster 3 = 24 items. **Nilai K = 5** Cluster 0: 48 items, Cluster 1: 37 items Cluster 2: 30 items, Cluster 3: 37 items, Cluster 4: 19 items. Dalam Pengambilan Keputusan Cluster terbaik menggunakan K-means akan dilihat melalui Nilai Davies Bouldin dengan perolehan rata-rata mendekati angka 0. Dari 4 Nilai K yang digunakan untuk clusterisasi, proses analisis Nilai K = 3 memiliki Nilai Davies Bouldin yang baik dari yang lain yaitu **0.968**. Nilai K= 3 memiliki performa distance yang baik sehingga cluster data ini yang akan digunakan untuk rekomendasi perangkan kondisi gizi pada anak, data tersebut dapat dilihat pada tabel 3.

Tabel 4 Cluster 0

No.	Cluster	Berat Badan	Tinggi Badan	Lingkar Kepala
1	Cluster_0	0.4	0.38	0.36
2	Cluster_0	0.32	0.10	0.54
3	Cluster_0	0.11	0.05	0.36
4	Cluster_0	0.20	0.31	0.29
5	Cluster_0	0.304	0.40	0.18
6	Cluster_0	0.37	0.20	0.36
7	Cluster_0	0.26	0.34	0.36
8	Cluster_0	0.26	0.27	0.545
9	Cluster_0	0.24	0.28	0.36
10	Cluster_0	0.13	0.06	0.43
...
53	Cluster_0	0.31	0.63	0.09

Tabel 5 Cluster 1

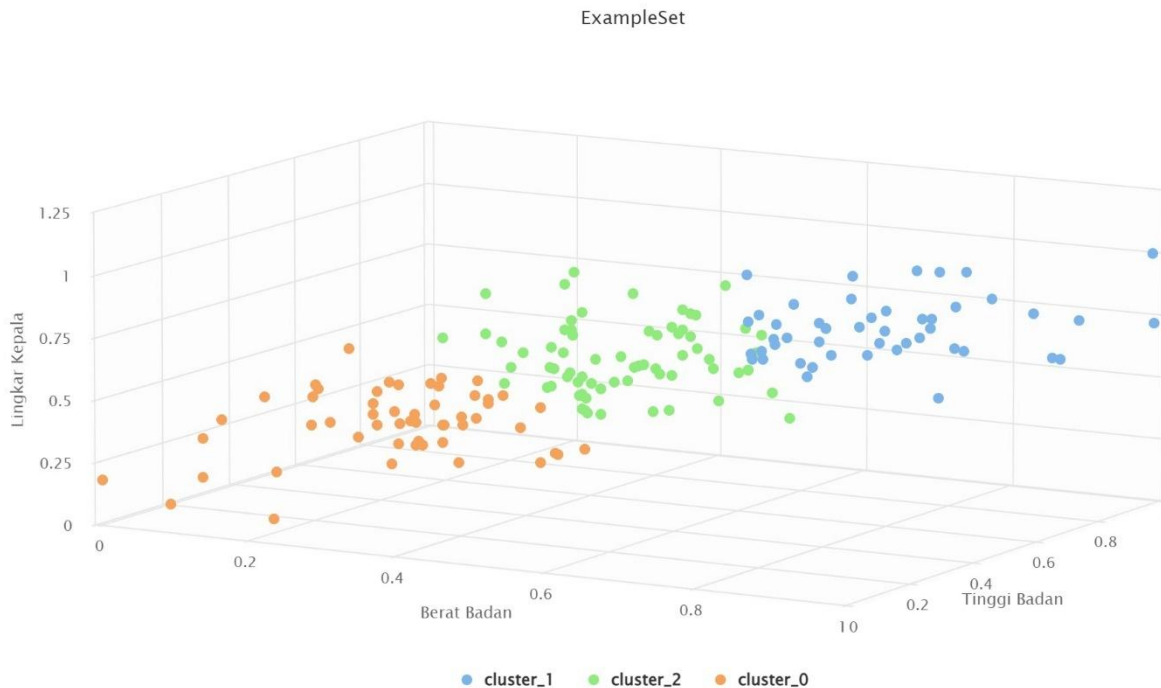
No.	Cluster	Berat Badan	Tinggi Badan	Lingkar Kepala
1	cluster_1	0.60	0.93	0.72
2	cluster_1	0.73	0.86	0.63
3	cluster_1	0.68	0.86	0.72
4	cluster_1	0.81	0.77	0.81
5	cluster_1	0.79	0.76	0.72
6	cluster_1	0.68	0.76	0.90

7	cluster_1	0.59	0.73	0.63
8	cluster_1	0.87	0.97	0.54
9	cluster_1	0.60	0.93	0.72
10	cluster_1	0.73	0.86	0.63
...
48	cluster_1	0.84	0.95	0.72

Tabel 6 Cluster 2

No.	Cluster	Berat Badan	Tinggi Badan	Lingkar Kepala
1	cluster_2	0.59	0.65	0.54
2	cluster_2	0.44	0.43	0.72
3	cluster_2	0.38	0.43	0.63
4	cluster_2	0.47	0.36	0.81
5	cluster_2	0.39	0.35	0.54
6	cluster_2	0.59	0.62	0.54
7	cluster_2	0.49	0.62	0.36
8	cluster_2	0.59	0.58	0.90
9	cluster_2	0.32	0.77	0.36
10	cluster_2	0.34	0.48	0.54
...
70	cluster_2	0.59	0.65	0.54

Berdasarkan ketiga tabel data cluster dan hasil analisa lanjutan cluster dilabeli dengan 3 perangkingan menjadi **Cluster 0** = Gizi Buruk , **Cluster 1** = Gizi Cukup , **Cluster 2** = Gizi Baik . Berikut Visualisasi data menggunakan Scatter 3D view.



Gambar 5. Visualisasi Data

KESIMPULAN

Berdasarkan Hasil penelitian perhitungan nilai akhir mahasiswa menggunakan K-means dapat disimpulkan bahwa penggunaan K-Means dapat dengan mudah mengelompokan data untuk dijadikan rekomendasi perangkingan, Hasil 3 cluster dengan nilai Davies Bouldin = 0.968 menggunakan nilai $K = 3$ menghasilkan rangking; Gizi buruk 53 anak, Gizi Cukup 48 anak , dan Gizi baik 70 anak dari total 171 anak yang dianalisis. Pengelompokan data menggunakan K-Means menjadi efektif jika dataset diatur dengan baik menggunakan besaran jarak angka yang tidak terlalu jauh sehingga memerlukan proses normalisasi data untuk mendapatkan hasil clustering yang maksimal.

DAFTAR PUSTAKA

- [1] C. A. Rahmat, H. Permatasari, E. Rasywir, and Y. Pratama, “Penerapan K-Means Untuk Clustering Kondisi Gizi Balita Pada Posyandu,” vol. 7, pp. 207–213, 2023, doi: 10.30865/mib.v7i1.5142.
- [2] D. Dona and M. Rifqi, “Penerapan Metode K-Means Clustering Untuk Menentukan Status Gizi Baik Dan Gizi Buruk Pada Balita (Studi Kasus Kabupaten Rokan Hulu),” *Rabit J. Teknol. dan Sist. Inf. Univrab*, vol. 7, no. 2, pp. 179–191, 2022, doi: 10.36341/rabit.v7i2.2171.
- [3] I. Maulana and U. Rosalina, “Clustering Data Nilai Ujian Akhir Semester Menggunakan Algoritma Data Mining K-Means,” *PERISKOP J. Sains dan Ilmu Pendidik.*, vol. 1, no. 2, pp. 76–85, 2020, doi: 10.58660/periskop.v1i2.10.
- [4] A. Harmain, H. Kurniawan, D. Maulina, and M. Universitas Amikom Yogyakarta, “Data Normalization for K-Means Efficiency on Groups of Areas With Potential Fores and /Land Fire Based on Heat Spots Distribution,” *Teknimedia*, vol. 2, no. 2, pp. 83–89, 2021.
- [5] M. P. A. Ariawan, I. B. A. Peling, and G. B. Subiksa, “Prediksi Nilai Akhir Matakuliah Mahasiswa Menggunakan Metode K-Means Clustering (Studi Kasus : Matakuliah Pemrograman Dasar),” *J. Nas. Teknol. dan Sist. Inf.*, vol. 9, no. 2, pp. 122–131, 2023, doi: 10.25077/teknosi.v9i2.2023.122-131.